The Dative Alternation in 20th century Newspaper Language

ICAME 33, Leuven 30 May - 2 June

Melanie Röthlisberger
University of Zurich
melanie.roethlisberger@uzh.ch

1 Background

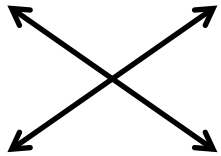
- Scope of the study:
  - analysis of Dative Alternation with more data points (1930/1960/1990)
    (see Grimm & Bresnan 2009)
  - Aims:
    - to account for and explore diachronic changes from 1930 to 1990 in Newspaper Language
    - to determine the impact of variables on the choice of dative constructions
    - to determine the interrelatedness between the variables and the relative importance of each

1 Background

- The Dative Alternation

The double-object construction

I	gave	Mary	the book.
		recipient	theme



The prepositional construction

I	gave	the book	to Mary.
		theme	recipient

2 Corpus data

- Brown-family of corpora including B-Brown and BLOB

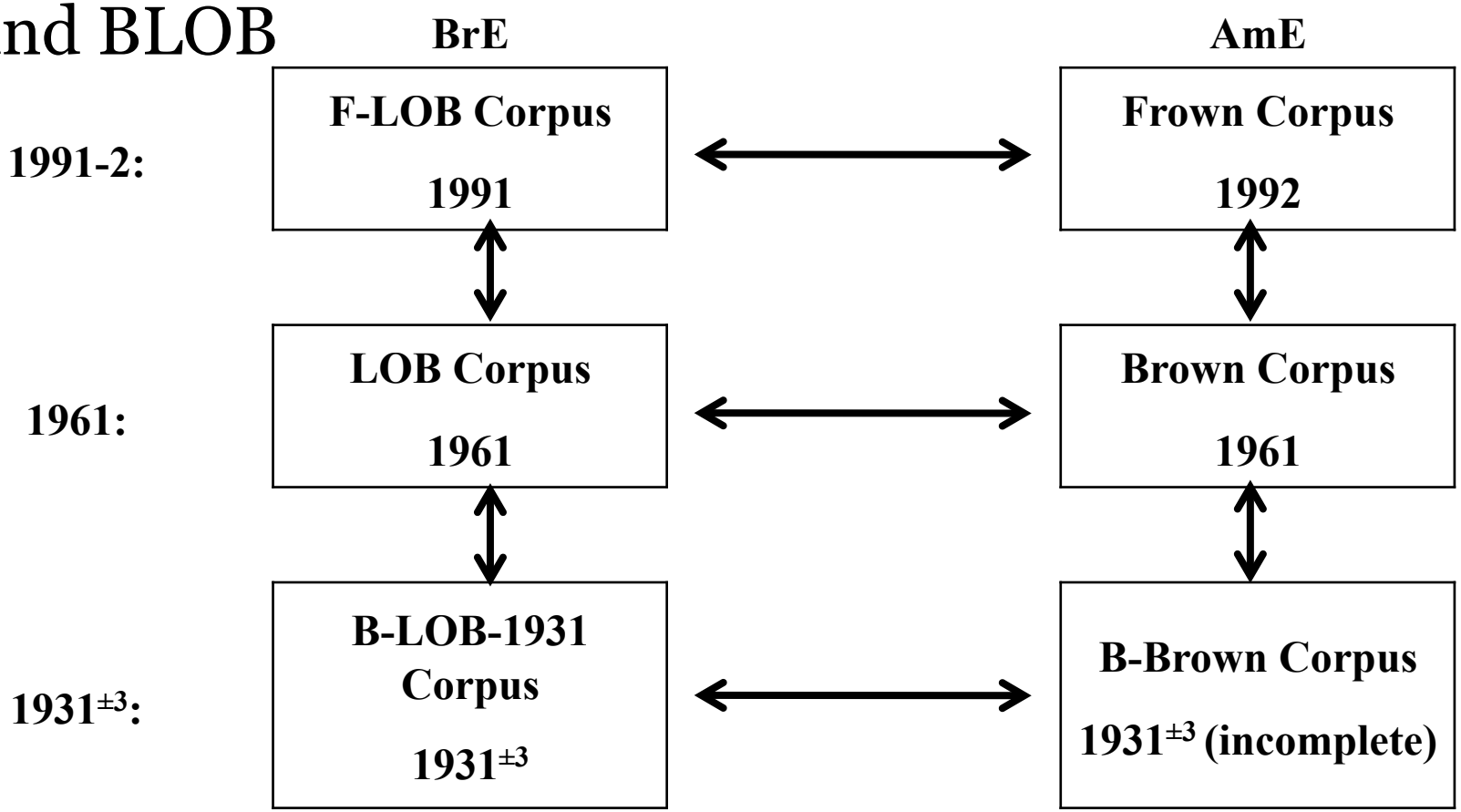


Figure 1. The Brown-family of corpora

2 Corpus data

Category	Text type	No. of files
A	Press Reportage	44 text files
B	Press Editorial	27 text files
C	Press Reviews	17 text files
D	Religion	17 text files
E	Skills and Hobbies	36 text files
F	Popular Lore	48 text files
G	Belles lettres, Memoirs	75 text files
H	Miscellaneous (Official documents)	30 text files
J	Scientific Writing	80 text files
K	General fiction	29 text files
L	Detective and Mystery Fiction	24 text files
M	Science Fiction	6 text files
N	Adventure and Western Fiction	29 text files
O	Romance and Love Story	29 text files
R	Humor	9 text files

Table 1. Text categories in the Brown-family of corpora

3 Extraction and filtering

Parser:

The data was parsed, using pro3gres, a dependency parser developed at the University of Zurich (Schneider 2008)

Extraction of [V NP NP] and [V NP to NP]:

The relevant constructions were extracted from the parsed data using a script written in prolog

3.1 Principles of exclusion

- alternation = always possible if semantically equivalent, grammatical correctness not considered (cf. Bresnan & Ford 2010: 171)
- no gerunds / phrasal constituents (!)
- no benefactives ("for")
- idiomatic expression and locatives included, as they allow in some instances for the alternation

  eg. "Stories like these must give people whose idea of heaven is a world without religion the creeps…"
  (example taken from Bresnan et al. 2007)

3.2 First results

	American English		**British English**	
	ditransitive	to-NP	ditransitive	to-NP
1930	72	148	68	144
1960	68	139	69	132
1990	63	140	77	162

Table 2. Raw numbers of double-object and prepositional construction in all six corpora.

Significance test, at df=2, X^2=0.19(AmE) and 0.30 (BrE)
→ the differences are NOT significant (for the factor TIME and VARIETY)

3.2 First results

	American English		**British English**	
	ditransitive	to-NP	ditransitive	to-NP
1930	33%	67%	32%	68%
1960	33%	67%	34%	66%
1990	31%	69%	32%	68%

Table 3. Percentages of double-object and prepositional constructions in all six corpora.

4 Coding of Variables

- Manual coding
- Length of recipient and theme: in characters
- Animacy of recipient and theme: animate vs inanimate
- Pronounhood of recipient and theme: pronoun vs NP
- Definiteness of recipient and theme: definite vs indefinite
- Giveness of recipient and theme: given vs new
- Period: 1930/1960/1990
- Variety: us/gb

5.1 Results: 'Length'

Mean length of Rec and Theme in AmE

	American English			
	ditransitive clauses		to-NP clauses	
	Rec	Theme	Rec	Theme
1930	6.2	23.4	22.9	14.6
1960	7.6	38.1	25.7	14.1
1990	7.1	25.1	24.8	13.9

Table 4. Mean length of Rec and Theme in American English.

Test of Significance: differences are NOT significant (ditransitive: df=2, X^2=0.413; to-NP: df=2, X^2 =0.118)

5.1 Results: 'Length'

Mean length of Rec and Theme in BrE

	British English			
	ditransitive clauses		to-NP clauses	
	Rec	Theme	Rec	Theme
1930	6.8	28.5	24.6	12.3
1960	7.1	26.8	25.5	12.7
1990	6.0	31.3	25.8	13.8

Table 5. Mean length of Rec and Theme in British English.

differences are NOT significant (p>0.20)

5.2 Results: Summary

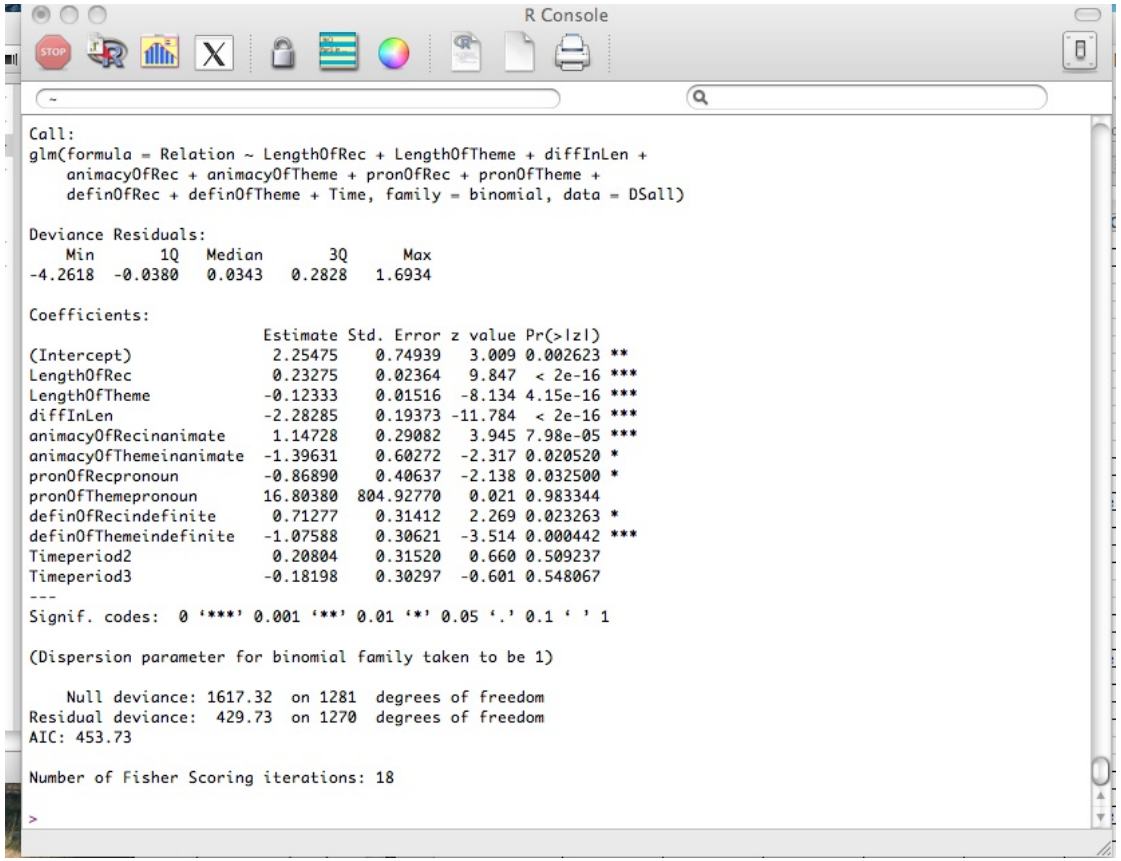
- Length: differences are not significant
- Animacy: differences are not significant
- Pronounhood: differences are not significant
- Definiteness: differences are not significant

(apart from: differences between defRec/defTH for double-object and prepositional dative construction are highly significant at X^2=19.759, df=6, p<0.01 in BrE)

5.3 Logistic regression analysis

- measures impact of independent variables on choice of constituents in relation to other variables

- analyses more than one independent variable simultaneously



Significance codes:
0: '***'
0.001: '**'
0.01: '*'
0.05: '.'
0.1: ' '

5.3 Logistic regression analysis

Regression analysis for all data					
Coefficients:	Estimate	Std. Error	z-value	Pr(>|z|)	
(Intercept)	2.25475	0.74939	3.009	0.002623	**
LengthOfRec	0.23275	0.02364	9.847	<2e-16	***
LengthOfTheme	-0.12333	0.01516	-8.134	<4.15e-16	***
diffInLen	-2.28285	0.19373	-11.784	<2e-16	***
animacyOfRec: inanimate	1.14728	0.29082	3.945	7.98E-05	***
animacyOfTheme: inanimate	-1.39631	0.60272	-2.317	0.02052	*
pronOfRec: pronoun	-0.8689	0.40637	-2.138	0.0325	*
pronOfTheme: pronoun	16.8038	804.9277	0.021	0.983344	
definOfRec: indefinite	0.71277	0.31412	2.269	0.023263	*
definOfTheme: indefinite	-1.07588	0.30621	-3.514	0.000442	***
Time: 1960	0.20804	0.3152	0.66	0.509237	
Time: 1990	-0.18198	0.30297	-0.601	0.548067	

5.3 Logistic regression analysis

Regression analysis for 1930-US					
Coefficients:	Estimate	Std. Error	z-value	Pr(>|z|)	
(Intercept)	0.39359	2.42857	0.162	0.871252	
LengthOfRec	0.42523	0.11684	3.639	0.000273	***
LengthOfTheme	-0.11468	0.04278	-2.68	0.007351	**
diffInLen	-2.67136	0.50983	-5.24	1.61E-07	***
animacyOfRec: inanimate	1.21791	0.79804	1.526	0.126977	
animacyOfTheme: inanimate	-1.7581	1.9364	-0.908	0.363919	
pronOfRec: pronoun	1.53703	1.433	1.073	0.283451	
pronOfTheme: pronoun	16.88028	1627.42746	0.01	0.991724	
definOfRec: indefinite	1.90869	0.8731	2.186	0.028808	*
definOfTheme: indefinite	-2.20647	0.86213	-2.559	0.010488	*

5.3 Logistic regression analysis

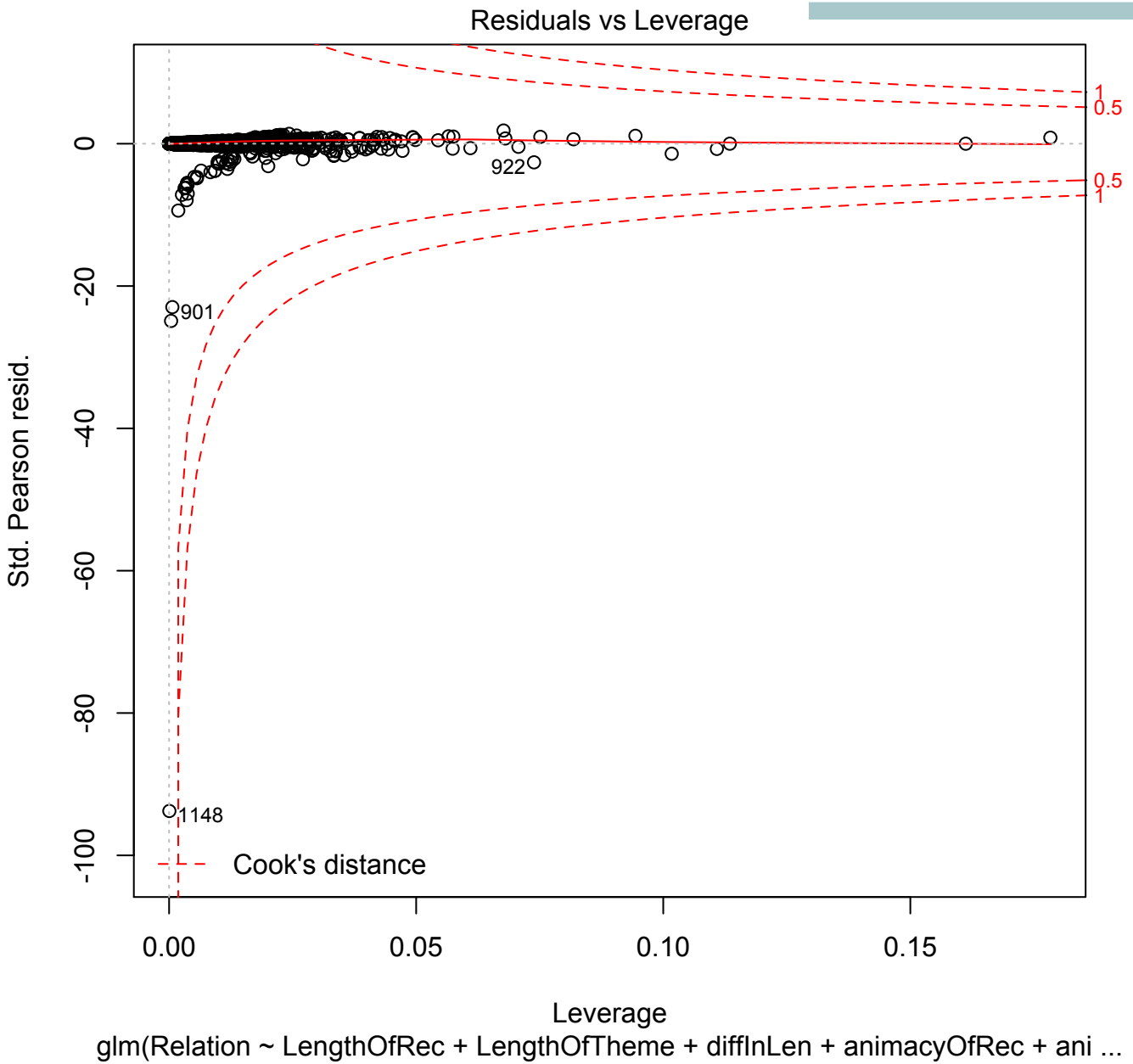
Regression analysis for 1930-GB					
Coefficients:					
	Estimate	Std. Error	z-value	Pr(>|z|)	
(Intercept)	16.24736	2426.53239	0.007	0.994658	
LengthOfRec	0.32606	0.0874	3.731	0.000191	***
LengthOfTheme	-0.16782	0.05381	-3.119	0.001816	**
diffInLen	-3.23549	0.84145	-3.845	0.00012	***
animacyOfRec: inanimate	2.70165	1.15513	2.339	0.019344	*
animacyOfTheme: inanimate	-16.67461	2426.53189	-0.007	0.994517	
pronOfRec: pronoun	-1.09661	1.26295	-0.868	0.385234	
pronOfTheme: pronoun	18.74695	2740.48867	0.007	0.994542	
definOfRec: indefinite	1.76432	1.08373	1.628	0.103524	
definOfTheme: indefinite	-0.11184	0.82807	-0.135	0.892559	

5.3 Logistic regression analysis

Regression analysis US-all times					
Coefficients:	Estimate	Std. Error	z-value	Pr(>|z|)	
(Intercept)	1.91805	1.10488	1.736	0.082567	.
LengthOfRec	0.22524	0.03339	6.746	1.52E-11	***
LengthOfTheme	-0.11183	0.02104	-5.315	1.07E-07	***
diffInLen	-2.28605	0.26621	-8.587	<2.00E-16	***
animacyOfRec: inanimate	1.31037	0.41542	3.154	0.001609	**
animacyOfTheme: inanimate	-1.07717	0.88642	-1.215	0.224289	
pronOfRec: pronoun	-0.67902	0.61549	-1.103	0.269932	
pronOfTheme: pronoun	16.62739	1194.50477	0.014	0.988894	
definOfRec: indefinite	1.1072	0.447	2.477	0.01325	*
definOfTheme: indefinite	-1.64839	0.46706	-3.529	0.000417	***
Timeperiod2	0.69975	0.46701	1.498	0.134044	
Timeperiod3	-0.09456	0.42899	-0.22	0.825537	

5.3 Logistic regression analysis

Regression analysis GB-all times					
Coefficients:	Estimate	Std. Error	z-value	Pr(>|z|)	
(Intercept)	2.46059	1.08341	2.271	0.0231	*
LengthOfRec	0.24984	0.03506	7.125	1.04E-12	***
LengthOfTheme	-0.1409	0.02408	-5.85	4.90E-09	***
diffInLen	-2.39658	0.30229	-7.928	2.23E-15	***
animacyOfRec: inanimate	1.10833	0.43376	2.555	0.0106	*
animacyOfTheme: inanimate	-1.62401	0.86618	-1.875	0.0608	.
pronOfRec: pronoun	-0.98203	0.56623	-1.734	0.0829	.
pronOfTheme: pronoun	17.019	1079.84065	0.016	0.9874	
definOfRec: indefinite	0.34537	0.46487	0.743	0.4575	
definOfTheme: indefinite	-0.67813	0.42683	-1.589	0.1121	
Timeperiod2	-0.09195	0.44687	-0.206	0.837	
Timeperiod3	-0.11991	0.44301	-0.271	0.7866	


Residuals vs Leverage
Std. Pearson resid.
0
-20
-40
-60
-80
-100
0.00
0.05
0.10
0.15
Leverage
1
0.5
0.5
1
922
901
1148
Cook's distance
glm(Relation ~ LengthOfRec + LengthOfTheme + diffInLen + animacyOfRec + ani ...

6 Tentative conclusion

- There are no significant changes between 1930, 1960 and 1990 with regard to the variables 'animacy', 'length', 'pronounhood' and 'definiteness'
- The mean length of the last constituent in both double-object and prepositional construction seems to increase over time in BrE and AmE (although not significant)
- Across all data:
  - The variable 'length' seems to be the most important variable (very highly significant) in all three periods and in both varieties
  - Animacy of recipient and definiteness of theme seem to be highly significant variables across all periods and both varieties
- Differences across the varieties:
  - While in AmE, the variable 'animacy of recipient' and 'definiteness of theme' are very highly significant variables
  - in BrE, animacy of recipient and theme, and pronounhood of theme are significant (at $p>0.05$ and $p>0.01$)

7 Outlook

- Provide verb list that occur in both constructions for every corpus for automatic extraction (choice context?)
- control for outliers and adapt model
- include more variables (eg. postmodification, persistence, thematicity, etc.)
- Interaction between variables

References

Arnold, Jennifer, Wasow, Thomas, Losongco, Anthony and Ryan Ginstrom. 2000. Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language* 76(1): 28-55.

Bresnan, Joan, Anna Cueni, Tatiana Nikitina, and Harald Baayen. 2007. Predicting the dative alternation. In: G. Boume, I. Kraemer, and J. Zwarts (eds.). *Cognitive Foundations of Interpretation*. Amsterdam: Royal Netherlands Academy of Science, 69-94.

Bresnan, Joan and Tatiana Nikitina. 2009. The Gradience of the Dative Alternation. In: Linda Uyechi and Lian-Hee Wee (eds.). *Reality Exploration and Discovery: Pattern Interaction in Language and Life*. Stanford: CSLI Publications, 1-23.

Bresnan, Joan and Marilyn Ford. 2010. Predicting Syntax: Processing Dative Constructions in American and Australian Varieties of English. *Language* 86(1):168-213.

Collins, Peter. 1995. The indirect object construction in English: an informational approach, *Linguistics* 33:35-49.

Gries. Stefan. 2003. *Multifactorial analysis in corpus linguistics: a study of Particle Placement*. London & New York: Continuum Press.

Grimm, Scott and Joan Bresnan. 2009. *Spatiotemporal variation in the dative alternation: A study of four corpora of British and American English*. Paper presented at the 3rd international conference on Grammar and Corpora, Mannheim, 22-24 September 2009.

Hawkins, John. 1994. *A performance theory of order and constituency*. Cambridge: Cambridge University Press.

Hawkins, John. 2000. The relative order of prepositional phrases in English: Going beyond Manner-Place-Time. *Language Variation and Change* 11: 231-266.

Hundt, Marianne and Geoffrey Leech. 2011. ‚Small is beautiful' – on the value of standard reference corpora for observing recent grammatical change. In: Terttu Nevalainen and Elizabeth Closs Traugott (eds.). *Handbook on the History of English: Rethinking Approaches to the History of English*. Oxford: Oxford University Press.

Preacher, K. J. (2001, April). Calculation for the chi-square test: An interactive calculation tool for chi-square tests of goodness of fit and independence [Computer software]. Available from http://quantpsy.org.

Rosenbach, Anette. 2002. *Genitive variation in English: conceptual factors in synchronic and diachronic studies*. Berlin: Mouton de Gruyter.

Szmrecsanyi, Benedikt and Lars Hinrichs. 2007. Recent changes in the function and frequency of Standard English genitive constructions: a multivariate analysis of tagged corpora. *English Language and Linguistics* 11(3): 437-474.

Schneider, Gerold. 2008. *Hybrid Long-Distance Functional Dependency Parsing*. *Doctoral Thesis*. Faculty of Arts, University of Zurich.

Wasow, Thomas. 1997. Remarks on grammatical weight. *Language Variation and Change* 9:81-105.

Wasow, Thomas and Jennifer Arnold. 2003. Post-verbal constituent ordering in English. In: Günter Rohdenburg and Britta Mondorf (eds.). *Determinants of Grammatical Variation in English*. Berlin/New York: Mouton de Gruyter, 119-154.

THANK YOU!

Coding of Variables

- Syntactic variable: length
  - measured in number characters, shown to be more stable than words
  - difference in length calculated as sign preserving log transform: $f(x) = \text{sgn}(x) \log(1+ |x|)$

- Semantic variable: animacy
  - four-scaled animacy hierarchy: human > animal > collective > inanimate
  - for this study: animate vs inanimate

Coding of Variables

- Morphosyntactic variable: pronounhood & definiteness
  - Pronounhood: full lexical NP vs pronoun substituting a lexical NP (Bresnan et al. 2007)
  - Definiteness: existential reading of constituent in "There is/are.." = indefinite; articles, demonstratives, personal and impersonal pronouns, wh pronouns, quantifiers, proper names = definite

- Discourse-functional variable: giveness
  - not measured yet
  - Outlook: constituent = given, if occurrence within the last 50 words of discourse / if second or first person pronoun (Hinrichs & Szmrecsanyi 2007)

5.1 Results: 'Length'

Difference in 'Length' as sign preserving log-value

	American English		**British English**	
	ditransitive	to-NP	ditransitive	to-NP
1930	1.03	0.33	1.10	0.58
1960	1.28	0.40	1.09	0.62
1990	0.97	0.40	1.16	0.39

Table 6. Sign preserving log value for difference in length.

Results: 'Animacy'

	American English			
	ditransitive clauses			
	animateRec	inanimateRec	animateTH	inanimateTH
1930	64	8	1	71
1960	54	14	3	65
1990	52	11	4	59

	American English			
	to-NP			
	animateRec	inanimateRec	animateTH	inanimateTH
1930	72	76	34	114
1960	60	79	29	110
1990	56	84	34	106

differences are NOT significant (also not for BrE)

Results: 'Animacy'

Percentages of animate recipients and themes

	American English				**British English**			
	ditransitive clauses		to-NP clauses		ditransitive clauses		to-NP clauses	
	Rec	Theme	Rec	Theme	Rec	Theme	Rec	Theme
1930	88.89%	1.39%	48.65%	22.97%	92.65%	0.00%	36.81%	22.22%
1960	79.41%	4.41%	43.17%	20.86%	84.06%	2.90%	52.27%	21.21%
1990	82.54%	6.35%	40.00%	24.29%	87.01%	2.60%	42.59%	25.31%

Results: 'Pronounhood'

	American English							
	ditransitive clauses				to-NP clauses			
	pronRec	nonpron Rec	pronTH	nonpron TH	pronRec	nonpron Rec	pronTH	nonpron TH
1930	41	31	0	72	6	142	26	122
1960	36	32	0	68	1	138	16	123
1990	36	27	0	63	5	135	17	123

	British English							
	ditransitive clauses				to-NP clauses			
	pronRec	nonpron Rec	pronTH	nonpron TH	pronRec	nonpron Rec	pronTH	nonpron TH
1930	42	26	0	68	8	136	22	122
1960	39	30	0	69	2	130	24	108
1990	42	35	0	77	10	152	25	137

Results: 'Pronounhood'

	American English			
	ditransitive clauses		to-NP clauses	
	Rec	Theme	Rec	Theme
1930	56.94%	0.00%	4.05%	17.57%
1960	52.94%	0.00%	0.72%	11.51%
1990	57.14%	0.00%	3.57%	12.14%

	British English			
	ditransitive clauses		to-NP clauses	
	Rec	Theme	Rec	Theme
1930	61.76%	0.00%	5.56%	15.28%
1960	56.52%	0.00%	1.52%	18.18%
1990	54.55%	0.00%	6.17%	15.43%

→ Changes not significant ($p>0.20$)

Results: 'Definiteness'

	American English							
	ditransitive clauses				to-NP clauses			
	defRec	indefRec	defTH	indefTH	defRec	indefRec	defTH	indefTH
1930	64	8	15	57	91	57	87	61
1960	61	7	17	51	94	45	61	78
1990	55	8	16	47	94	46	64	76

Changes are not significant ($p>0.20$)

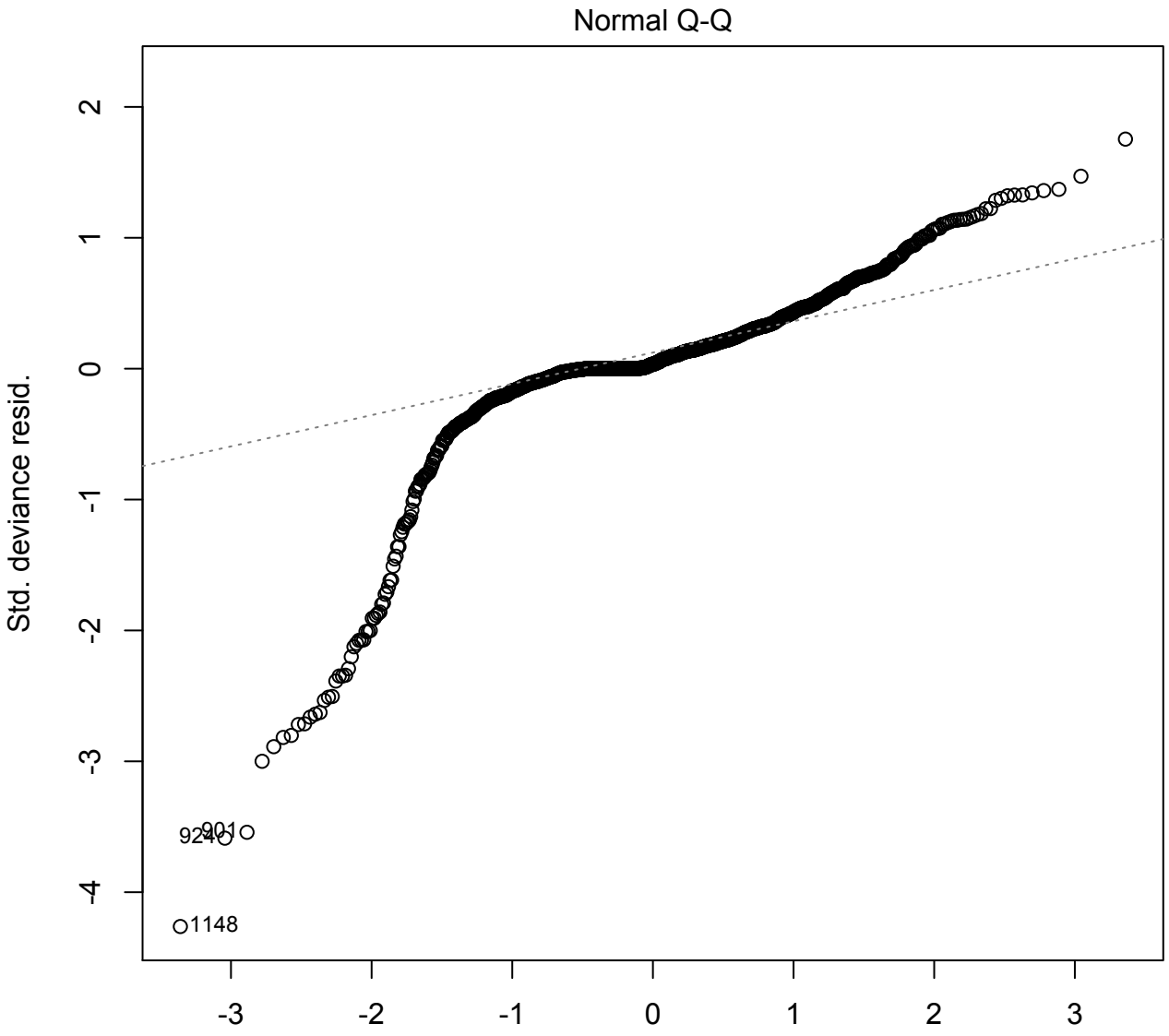
Results: 'Definiteness'

	British English							
	ditransitive clauses				to-NP clauses			
	defRec	indefRec	defTH	indefTH	defRec	indefRec	defTH	indefTH
1930	64	4	22	46	103	41	54	90
1960	60	9	23	46	99	33	64	68
1990	63	14	25	52	106	56	78	84

- differences between defRec/defTH for double-object and prepositional dative construction are highly significant at $X^2=19.759$, df=6, $p<0.01$; not so in American English

Results: 'Definiteness'

	American English				**British English**			
	ditransitive clauses		to-NP clauses		ditransitive clauses		to-NP clauses	
	Rec	Theme	Rec	Theme	Rec	Theme	Rec	Theme
1930	88.89%	20.83%	61.49%	58.78%	94.12%	32.35%	71.53%	37.50%
1960	89.71%	25.00%	67.63%	43.88%	86.96%	33.33%	75.00%	48.48%
1990	87.30%	25.40%	67.14%	45.71%	81.82%	32.47%	65.43%	48.15%


Normal Q-Q
Std. deviance resid.
2
1
0
-1
-2
-3
-4
9240
901
1148
-3
-2
-1
0
1
2
3
Theoretical Quantiles
glm(Relation ~ LengthOfRec + LengthOfTheme + diffInLen + animacyOfRec + ani ...